Available online at www.sciencedirect.com

ScienceDirect

Molecular and Cellular Probes 21 (2007) 295–302

Molecular and Cellular Probes

www.elsevier.com/locate/ymcpr

Sequencing and analysis of the *Escherichia coli* serogroup O117, O126, and O146 O-antigen gene clusters and development of PCR assays targeting serogroup O117-, O126-, and O146-specific DNA sequences

Yanhong Liu[a], Chitrita DebRoy[b], Pina Fratamico[a,*]

[a]*Eastern Regional Research Center, US Department of Agriculture, Agricultural Research Service, Wyndmoor, PA 19038, USA*
[b]*Gastroenteric Disease Center, Department of Veterinary Science, The Pennsylvania State University, University Park, PA 16802, USA*

Received 30 October 2006; accepted 8 March 2007
Available online 14 March 2007

Abstract

The O-antigen gene clusters of *Escherichia coli* serogroups O117, O126, and O146 were sequenced, and 11, 10, and 11 open reading frames (ORFs) were identified, respectively. Genes required for O-antigen sugar biosynthesis, sugar transfer, and sugar processing were identified. Multiplex polymerase chain reaction (PCR) assays were developed targeting the *wzx* and *wzy* genes present in the O-antigen gene cluster of these serogroups. The assays were highly serogroup specific when tested against strains belonging to serogroups that were isolated from food, humans, animals, and environmental sources, as well as against representative strains belonging to ca. 165 different *E. coli* O serogroups and a number of non-*E. coli* bacteria. Thus, the results demonstrate that the *wzx* and *wzy* gene sequences were specific to *E. coli* O117, O126, and O146 and can be used as diagnostic markers for rapid identification and detection of these serogroups.
Published by Elsevier Ltd.

Keywords: *wzx*; *wzy*; Serotyping; *E. coli* O117; *E. coli* O126; *E. coli* O146; Serogroup; O-antigen; PCR

1. Introduction

Escherichia coli, which includes both commensal and pathogenic strains, is the most thoroughly studied bacterial species in the microbial world. The six major categories of *E. coli* strains that can cause enteric diseases in humans are (1) enterohemorrhagic *E. coli* (EHEC), (2) enterotoxigenic *E. coli* (ETEC), (3) enteropathogenic *E. coli* (EPEC), (4) enteroaggregative *E. coli* (EAggEC), (5) enteroinvasive *E. coli* (EIEC), and (6) diffusely adherent *E. coli* (DAEC) [1].

ETEC strains are a major cause of traveler's diarrhea. This disease is characterized by adherence to and colonization of the intestinal mucosa by the bacteria and the production of the enterotoxins, heat-labile toxin (LT) and/or heat-stable toxin (ST). Strains of *E. coli* serogroup O117 have been identified as ETEC due to the presence of heat-stable enterotoxin genes and their ability to adhere to the brush border of human enterocytes [2]. *E. coli* O117 strains have been associated with acute diarrhea in infants in Africa [3]. In addition to enteric disease, *E. coli* O117 was also implicated in urinary tract infection through sexual transmission [4].

EPEC strains are a major cause of diarrhea in children living in developing countries and are characterized by the presence of two virulence markers: a chromosomal *eae* gene involved in intimate attachment to intestinal cells and a plasmid harboring genes involved in localized adherence. *E. coli* serogroup O126 is a class I EPEC [5], even though some strains of this serogroup do not possess the virulence markers described above [6]. *E. coli* O126 strains have been associated with sporadic cases and outbreaks of infantile diarrhea [6]. Although strains of *E. coli* O126 are human pathogens, they are also found in healthy animals, such as dairy goats [7].

All EHEC strains produce Shiga toxins (Stx), also known as verotoxins or verocytotoxins. Thus, EHEC strains are also called Shiga toxin-producing *E. coli* (STEC) [5]. STEC are recognized as important pathogens and have

*Corresponding author. Tel.: +1 215 233 6525; fax: +1 215 233 6581.
E-mail address: pina.fratamico@ars.usda.gov (P. Fratamico).

0890-8508/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.mcp.2007.03.002

been associated with diarrhea, hemorrhagic colitis, and hemolytic uremic syndrome [8]. *E. coli* O146:H21 strains belonging to the STEC group have been involved in human disease [9] and have also been found in healthy animals, including sheep and cattle [10–12]. *E. coli* O146:H21 strains that elaborate Shiga toxins (Stx1 and Stx2) and that possess virulence genes, including *eae* and *ehxA,* have been identified [11–13].

E. coli serotyping is typically performed by agglutination reactions using antisera raised in rabbits against the ca. 165 different O standard references strains. However, traditional serotyping is both laborious and time consuming, and it often generates equivocal results due to cross-reactions between different serogroups. Furthermore, the antisera used for serotyping can only be generated by specialized laboratories with animal facilities. As such, rapid, more specific molecular methods for identifying different *E. coli* serogroups are needed.

The O-antigen, which contains many repeats of an oligosaccharide unit (O unit), is present in the outer membrane of Gram-negative bacteria and contributes the major antigenic variability to the cell surface. The genes involved in the biosynthesis of O-antigens in *E. coli* are located in the O-antigen gene cluster and are flanked by the *galF* and *gnd* genes on the *E. coli* chromosome. The genes that encode proteins within the *E. coli* O-antigen gene clusters consist of the following three categories: nucleotide sugar biosynthesis, glycosyltransferase, and O-antigen processing genes [14]. Nucleotide sugar biosynthesis gene products are involved in the biosynthesis of the nucleotide sugar precursors in the O-antigen, which occurs in the cytoplasm. Glycosyltransferases, usually have narrow substrate specificity and are responsible for transferring the various precursor sugars to form an oligosaccharide on a carrier lipid, undecaprenyl phosphate (UndP), which is located on the inner membrane facing the cytoplasmic side. O-antigen processing proteins include a flippase (Wzx) and the O-antigen polymerase (Wzy). The Wzx flips the O-unit across the inner membrane. After the UndPP-linked O-unit is translocated across the cytoplasmic membrane, the O-units are linked together by Wzy through a glycosidic linkage. Although both Wzx and Wzy are membrane proteins usually with high variation among different microorganisms, the action of Wzx is not very specific while the function of Wzy is specific [14].

A number of *E. coli* O-antigen gene clusters have been sequenced and the genes were annotated [8,15–21]. Several genes in the O-antigen gene clusters, in particular, the *wzx* (O-antigen flippase) and *wzy* (O-antigen polymerase) genes, show relatively low similarity among different *E. coli* serogroups, and PCR primers targeting the *wzx* and *wzy* genes have been used to develop serogroup-specific PCR assays [15–21]. Our objective was to sequence and characterize the O-antigen gene clusters of *E. coli* O117, O126, and O146 serogroups and identify specific genes that can be used as diagnostic markers for these serogoups.

2. Materials and methods

2.1. Bacterial strains

E. coli O117:H7 strain 97-3039, *E. coli* O126:H27 strain 89-3506, and *E. coli* O146:H21 strain 90-3158 used for DNA sequencing of the O-antigen gene clusters were obtained from The Centers for Disease Control and Prevention (CDC), Atlanta, GA. Bacteria used to test for specificity of the PCR included the following: 90 *E. coli* O117 strains, 77 *E. coli* O126 strains, 98 *E. coli* O146 strains, 47 non-O117, non-O126, and non-O146 *E. coli* strains isolated from humans, animals, food, and water, and 165 *E. coli* reference standard strains belonging to serogroups O1–O175 used for serotyping, except for O14, O31, O47, O67, O72, O93, O94 and O122 strains that have not been designated [22]. In addition, strains representative of other bacterial genera ($n = 20$) used to test the PCR specificity included *Bacillus cereus, Citrobacter freundii, Enterobacter cloacae, Enterococcus aerogenes, Enterococcus faecalis, Hafnia alvei, Klebsiella pneumonia, Listeria monocytogenes, Pseudomonas aeroginosa, Proteus vulgaris, Salmonella* Anatum, *Salmonella* Arizona, *Salmonella* Choleraesuis, *Salmonella* Enteritidis, *Salmonella* Typhimurium, *Serratia marcescens, Shigella boydii, Staphylococcus aureus, Vibrio cholerae, and Yersinia enterocolitiica*. All bacteria were grown in Luria Bertani (LB) broth or on LB agar plates at 37 °C.

2.2. Construction of a DNaseI shotgun library, DNA sequencing, and gene annotation

Genomic DNA was isolated using the DNeasy Tissue Kit (Qiagen, Inc., Valencia, CA) according to the manufacturer's instructions. Long PCR assays were performed to amplify the O-antigen gene clusters using the Expand Long Template PCR System (Roche Applied Science, Mannheim, Germany) and JUMPSTART (named for Just Upstream of Many Polysaccharide-associated gene STARTs) and GND (6-phosphogluconate dehydrogenase gene) primers that flank the *E. coli* O-antigen gene clusters. The sequence of the JUMPSTART sense primer was 5′-ATTGGTAGCTGTAAGCCAAGGGCGGTAGCGT-3′, and the antisense GND primer sequence was 5′-CACTGCCATACCGACGACGCCGATCTGTTGCTTGG-3′ (Invitrogen Life Technologies, Inc., Carlsbad, CA). The long PCR conditions were as described previously [17]. The long PCR products were verified on 0.8% agarose gels and purified according to the instructions in the QIAquick PCR Purification Kit (Qiagen, Inc., Valencia, CA). DNase I digestion, shotgun cloning, and DNA sequencing were performed as described previously [17]. To confirm the sequence of each O-antigen gene cluster, ten individual long PCR products were pooled together and the DNA was resequenced using primers designed from different regions of the O-antigen gene clusters. Sequencing data were assembled using Sequencher software, and gene annotation was performed as described previously [17]. The HMMTOP program [23] was used to identify potential transmembrane segments from the amino acid sequences.

2.3. Testing for specificity by the PCR

Bacterial DNA used as template for the PCR assays was isolated as described previously [17]. PCR primers were designed from the *wzx* and *wzy* genes found in the O-antigen gene clusters of each serogroup using the Primer3 software program. Multiplex PCR reactions were performed and analyzed as described [17].

2.4. Nucleotide sequence accession numbers

The DNA sequences of the O-antigen gene clusters of *E. coli* O117, O126, and O146 were deposited into GenBank under the accession numbers DQ465247, DQ465248, and DQ465249, respectively.

3. Results and discussion

DNA sequences of 10,886, 11,783, and 11,888 bases were obtained from the *E. coli* O117, O126, and O146 O-antigen gene clusters, respectively (Tables 1–3). All of the genes in each of the clusters had the same transcriptional direction from *galF* to *gnd.* The deduced amino acid sequences from these ORFs were used to search the NCBI database for indication of their possible functions. Gene names were assigned on the basis of the bacterial polysaccharide gene nomenclature system (http://www.microbio.usyd.edu.au/BPGD/big-paper.pdf).

3.1. Sequence analysis of the E. coli O117 O-antigen gene cluster

3.1.1. Sugar biosynthetic pathway genes

The structure of the O-unit of *E. coli* O117 has been characterized, and it contains four residues: *N*-acetyl galactosamine (GalpNAc), L-Rhamnose (L-Pha), glucose (Glcp), and galactose (Galp) [24] (Fig. 1). Since genes for the synthesis of common sugars, including glucose and galactose are located outside the O-antigen gene cluster, only genes involved in the synthesis of L-Rhamnose were expected in the *E. coli* O117 O-antigen gene cluster.

As shown in Table 1, ORF1, ORF2, ORF3, and ORF4 of the O117 O-antigen gene cluster showed between 77% and 98% identity to known Rml B, D, A, and C proteins, respectively. L-Rhamnose synthesis requires *rml* genes and these genes have been well characterized [14]. Recently, L-Rhamnose was synthesized in vitro using the enzymes found in the L-Rhamnose synthetic pathway [25]. Therefore, ORF1-4 were identified as *rmlB*, *rmlD*, *rmlA,* and *rmlC*, respectively, and named accordingly. The *rml* genes are usually found together at the 5′ end of the O-antigen gene clusters in the order of *rml*BDAC [14]. Our results are consistent with these findings; However, there are exceptions. For example, in the *E. coli* O177 and *Shigella boydii* type 9 O-antigen gene clusters the *rmlC* gene is separated from other *rml* genes by *wzx* [15,20].

ORF11 shows 80% similarity to a putative UDP-galactose-4-epimerase found in *E. coli*. The UDP-galactose-4-epimerase is the product of *galE,* and it catalyzes the conversion of UDP galactose to UDP glucose and the reverse reaction. Even though the *galE* gene is a housekeeping gene that is located elsewhere in the chromosome

Table 1
Open reading frames (ORFs) in the O-antigen gene cluster of *E. coli* serogroup O117

ORF #	Proposed gene name	Location	No. of amino acids	Putative function	Most significant homology (accession no.)	% Identity/% similarity
1	*rmlB*	300–1193	297	dTDP-glucose 4,6-dehydratase	dTDP-glucose 4,6-dehydratase [*E. coli*] (AAN60454.1)	98/98
2	*rmlD*	1193–2092	299	dTDP-6-deoxy-D-glucose-3,5 epimerase	dTDP-6-deoxy-D-glucose-3,5 epimerase [*E. coli*] (AAZ85715.1)	98/99
3	*rmlA*	2150–3028	292	Glucose-1-phosphate thymidylyltransferase	Glucose-1-phosphate thymidylyltransferase [*Shigella boydii*] (AAL27324.1)	99/100
4	*rmlC*	3033–3575	180	dTDP-6-deoxy-D-glucose-3,5 epimerase	Putative protein [*E. coli*] (AAZ65834.1)	77/86
5	wzx	3575–4792	405	O-antigen flippase	Polysaccharide biosynthesis protein [*Shewanella baltica* OS155] (ZP_00581317.1)	26/45
6	*wzy*	4785–6107	440	O-antigen polymerase	Secreted polysaccharide polymerase [*Bacillus cereus* ATCC 14579] (AAP12132.1)	22/40
7	*wbeA*	6100–6900	266	Glycosyl transferase, family 2	Glycosyl transferase, family 2 [*Pelobacter propionicus* DSM 2379] (ZP_00676473)	42/58
8	*wbeB*	6887–7831	314	Glycosyl transferase, family 2	Glycosyl transferase, family 2 [*Shewanella putrefaciens* CN-32] (ZP_00813359.1)	35/55
9	*wbeC*	7828–8889	353	Glycosyltransferase	Putative LPS biosynthesis related glycosyltransferase [*Bacteroides fragilis* NCTC 9343] (CAH09146.1)	33/51
10	*wbeD*	8886–9695	269	Glycosyl transferase	WclG [*E. coli*] (AAV74552.1)	50/67
11	*galE*	9709–10728	339	UDP-galactose-4-epimerase	Putative UDP-galactose-4-epimerase [*E. coli*] (AAS73174.1)	69/80

Table 2
Open reading frames (ORFs) located in the *E. coli* O126 O-antigen gene cluster

ORF #	Proposed gene name	Location	No. of amino acids	Putative function	Most significant homology (accession no.)	% Identity/% similarity
1	*wzy*	43–1086	347	O-antigen polymerase	Wzy [*Yersinia enterocolitica* (type 0:8)] (AAC60768.1)	34/57
2	*wbgK*	1083–2084	333	Glycosyltransferase	WbcF [*Yersinia enterocolitica* (type 0:8)] (AAC60769.1)	42/65
3	*wbgL*	2094–2987	297	Glycosyltransferase	glycosyltransferase [*Salmonella enterica* subsp. salamae serovar Greenside] (AAV34522.1)	42/57
4	*wbsT*	2994–3764	256	Glycosyltransferase	WblT protein [*Photorhabdus luminescens* subsp. laumondii TTO1] (CAE17191.1)	55/73
5	*wbsU*	3768–4898	376	Glycosyltransferase	WblU protein [*Photorhabdus luminescens* subsp. laumondii TTO1] (CAE17192.1)	39/60
6	*gmd*	4972–6093	373	GDP-D-mannose dehydratase	COG1089: GDP-D-mannose dehydratase [*Yersinia mollaretii* ATCC 43969]	81/90
7	*fcl*	6059	334	GDP-L-fucose synthetase	GDP-L-fucose synthetase [*Yersinia pseudotuberculosis* (type O:1b)] (CAB63301.1)	67/80
8	*manC*	7150–8547	465	Mannose-1-phosphate guanylyltransferase	Mannose-1-phosphate guanylyltransferase [*Aeromonas hydrophila*] (AAM74484.1)	56/75
9	*manB*	8611–10050	479	Phosphomannomutase	ManB [*E. coli*] (AAV74380.1)	54/73
10	*wzx*	10154–11638	494	Flippase	Polysaccharide transport protein, putative [*Bacillus cereus* ATCC 10987] (AAS44286.1)	30/53

Table 3
Open reading frames (ORFs) in the O-antigen gene cluster of *E. coli* serogroup O146

ORF	Proposed gene name	Location	No. of amino acids	Putative function	Most significant homology (accession no.)	% Identity/ % similarity
1	*rmlB*	301–1194	297	dTDP-glucose 4,6-dehydratase	dTDP-glucose 4,6-dehydratase [*E. coli*] (AAZ85703.1)	98/99
2	*rmlD*	1194–2093	299	dTDP-4-dyhydrorhamnose reductase	Putative dTDP-4-dyhydrorhamnose reductase [*E. coli*] (AAZ65832.1)	97/98
3	*rmlA*	2151–3029	292	Glucose-1-phosphate thymidylyltransferase	Glucose-1-phosphate thymidylyltransferase [*E. coli*] (AAZ85705.1)	98/100
4	*rmlC*	3034–3567	177	dTDP-6-deoxy-D-glucose-3,5 epimerase	dTDP-6-deoxy-D-glucose-3,5 epimerase [*E. coli*] (AAZ85706.1)	66/80
5	*wbuD*	3582–4574	330	Nitroreductase	Nitroreductase [*Methylobacillus flagellatus* KT] (EAN02140.1)	29/47
6	*wzx*	4571–6109	512	Flippase	Hypothetical protein BT_0040 [*Bacteroides thetaiotaomicron* VPI-5482] (AAO75147.1)	29/50
7	*wbuE*	6102–7205	367	Unknown	conserved hypothetical protein [*Bacteroides fragilis* NCTC 9343] (CAH08488.1)	32/50
8	*wbwU*	7552–8352	266	Glycosyl transferase	Eps4L [*Streptococcus thermophilus*] (AAN63688.1)	29/52
9	*wzy*	8472–9650	392	Polymerase	putative polysaccharide polymerase [*Staphylococcus saprophyticus* subsp. saprophyticus ATCC 15305] (BAE17214.1)	21/40
10	*wbwV*	9613–10,500	295	Glycosyltransferase	Glycosyl transferase, family 2 [*Pseudoalteromonas atlantica* T6c] (ZP_00776237.1)	46/62
11	*wbwW*	10,550–11,362	270	Glycosyltransferase	WbcZ [*E. coli*] (AAY28251.1)	51/67

a

3)-α-D-Gal*p*NAc-(1→4)-β-D-Gal*p*NAc-(1→3)-α-L-Rha*p*-(1→4)-α-D-Glc*p*-(1→4)-β-D-Gal*p*-(1→

b

→2)-β-D-Man*p*-(1→3)-β-D-Gal*p*-(1→3)-α-D-GlcNAc-(1→3)-β-D-Glc*p*NAc-(1→

α-L-Fuc*p*-(1→2)

Fig. 1. The structures of the *E. coli* O117 (a) [24] and O126 (b) [29] O-antigen oligosaccharides.

[14], this gene has been found in several O-antigen gene clusters, including O103 and O113 [8,26]. Experimentally, UDP-galactose-4-epimerase converts galactose to glucose in *E. coli* [27]. The *galE* gene is also critical in the synthesis of the side chains of bacterial lipopolysaccharides, since the absence of this gene product abolishes the synthesis of the O-specific side chain [28].

3.1.2. Sugar transferase genes

Glycosyltransferases are specific to different sugar donors, acceptors, and linkages between two sugars. Since the structure of the *E. coli* O117 O-unit is a pentasaccharide (Fig. 1), four glycosyltransferases were expected in the O-antigen gene cluster. In fact, ORF 7 and 8 share 58% and 55% similarity to proteins in the glycosyltransferase family 2 of *Pelobacter propionicus* DSM 2379 and *Shewanella putrefaciens* CN-32, respectively. Therefore, ORF7 and ORF8 were proposed to be gylcosyltransferase genes and named *wbeA* and *wbeB*, respectively. Likewise, ORF9 and ORF10 share 51% and 67% similarity with putative glycosyltransferases in *Bacteroides fragilis* NCTC 9343 and WclG in *E. coli*, respectively. Based on high levels of similarity, ORF9 and ORF10 were proposed to be glycosyltransferases and were named *wbeC* and *wbeD*, respectively.

3.1.3. O-antigen processing genes

The ORF5 shows low homology with a number of Wzx proteins that have been identified previously, and ORF6 shows 40% similarity to a secreted polysaccharide polymerase in *B. cereus*. Thus, ORF5 and ORF6 were named *wzx* and *wzy*, respectively. As expected, both Wzx and Wzy proteins were hydrophobic and contained 12 and 13 transmembrane domains, respectively, as analyzed by the HMMTOP program [23]. Since these genes are not very conserved among different *E. coli* serogroups, *wzx* and *wzy* are suitable targets for serogroup-specific PCR assay development.

3.2. Sequence analysis of the E. coli O126 O-antigen gene cluster

3.2.1. Sugar biosynthetic pathway genes

The structure of the *E. coli* O126 O-antigen contains the following pentasaccharide repeats: mannose, galactose, *N*-acetylglucosamine (two), and L-fucose [29] (Fig. 1). Since galactose and *N*-acetylglucosamine are synthesized elsewhere in the genome, only genes responsible for mannose and L-fucose are expected to be present in the O-antigen gene cluster. ORF6 shows 90% similarity to GDP-D-mannose dehydratase in *Yersinia mollaretii*, and ORF7 shows 80% similarity to GDP-L-fucose synthetase in *Yersinia pseudotuberculosis* (Table 2). Both GDP-mannose dehydratase (*gmd*) and GDP-L-fucose (*fcl*) were involved in the biosynthesis of L-fucose [18,19]. Therefore, ORF6 and ORF7 were named *gmd* and *fcl*, respectively. ORF8 displays 75% similarity to mannose-1-phosphate guanylyltransferase (ManC), and ORF9 shows 73% similarity to phosphomannomutase (ManB) in *E. coli*. Based on these similarities, ORF6 and ORF7 were named *manC* and *manB*, respectively. Both ManC and ManB have been well characterized and were involved in the biosynthesis of GDP-mannose [14]. The *gmd*, *fcl*, *manC,* and *manB* genes are also present in a number of other O-antigen gene clusters, including those of *E. coli* serogroups O86, O111, and O128 [16,18,19].

3.2.2. Sugar transferase genes

Since the O-unit of *E. coli* O126 has a pentasaccharide structure (Fig. 1), four glycosyltransferases were expected to be present in the O-antigen gene cluster of this *E. coli* serogroup. Consistent with this finding, ORF2-5 show different levels of similarities to glycosytransferases; therefore, they were proposed to have glycosyltransferase activity and named *wbgK*, *wbgL*, *wbsT,* and *wbsU*, respectively. As for other O-antigen genes, ORF2-5 may be responsible for linking the sugars together to form a pentasaccharide.

3.2.3. O-antigen processing genes

The O-antigen processing genes, *wzx* and *wzy*, are also present in the O-antigen gene cluster of *E. coli* O126, although their locations are different than in *E. coli* O117. ORF1 shows 57% similarity to Wzy in *Y. enterocolitica* (Table 2), and in addition, ORF1 has 10 transmembrane helices as revealed by the HMMTOP program. Therefore, ORF1 was named *wzy*. ORF10 shows 53% similarity to a putative polysaccharide transport protein in *Bacillus cereus* (Table 2) and also shows high homology to a number of flippases (data not shown). The HMMTOP program revealed that ORF10 has 14 transmembrane helices. Based on these results, ORF10 was expected to encode a flippase and named *wzx*, accordingly.

3.3. Sequence analysis of the E. coli O146 O-antigen gene cluster

3.3.1. Sugar biosynthetic pathway genes

The structure of the *E. coli* O146 polysaccharide has not yet been determined. However, ORF1, ORF2, ORF3, and ORF4 showed between 66% and 98% identity to known Rml B, D, A, and C proteins, respectively. This indicates that the *E. coli* O146 polysaccharide may contain Rhamnose similar to *E. coli* O117. ORF5 shows 47% similarity to a nitroreductase in *Methylobacillus flagellatus KT*. This gene has never been reported to be involved in O-antigen polysaccharide biosynthesis. ORF7 shows 50% similarity to an unknown protein. The presence of the unknown protein and the nitroreductase in the O-antigen gene cluster may indicate a novel polysaccharide structure in the O-unit of *E. coli* O146.

3.3.2. O-antigen processing genes

ORF6 shows 50% similarity to a hypothetical membrane protein in *Bacteroides thetaiotaomicron* speculated to be involved in the export of O-antigen and teichoic acid.

Analysis using HMMTOP revealed that ORF6 contains 14 transmembrane helices. In addition, ORF6 also shows homology to a number of flippases in other bacteria such as *Vibrio parahaemolyticus, B. fragilis*, and *Streptococcus pneumoniae*. Therefore, ORF6 was speculated to encode a flippase and named *wzx*. Since ORF9 shows 40% similarity to a putative polysaccharide polymerase in *Staphylococcus saprophyticus*, ORF9 was predicted to encode a polymerase and was named *wzy*. ORF9 also has 11 transmembrane helices as indicated by HMMTOP analysis.

3.3.3. Sugar transferase genes

ORF8 shows 52% similarity to Eps4L in *Streptococcus thermophilus*, a glycosyltransferase in group 1. ORF9 shows 62% similarity to a putative glycosyltransferase in family 2 in *Pseudoalteromonas atlantica* strain *T6c*. ORF11 shows 67% similarity to WbcZ in *E. coli*, which is a putative glycosyltransferase. Based on the homologies to glycosyltransferases, ORF8, 10, and 11 were proposed to be glycosyltransferases and named *wbwU*, *wbwV,* and *wbwW*, respectively. The fact that three glycosyltransferases are present in the O-antigen gene cluster indicates that the polysaccharide unit of *E. coli* O146 may be tetrasaccharide.

3.4. Identification of E. coli O117, O126, and O146 serogroup-specific genes and specificity testing

Sequencing analysis revealed that the *wzx* and *wzy* genes of the O-antigen gene clusters of *E. coli* O117, O126, and O146 share the least similarities with known *wzx* and *wzy* genes, and this is consistent with previous studies that demonstrated that the *wzx* and *wzy* genes were serogroup specific. Primers targeting the *E. coli* O117, O126, and O146 *wzx* and *wzy* genes were designed (Table 4) and used in multiplex PCR assays to determine their specificities for these serogroups. Multiplex PCR assays were performed, and the results are shown in Fig. 2. In this study, PCR reactions were performed to test the specificities against the ca. 165 *E. coli* standard strains, 47 strains of non-O117, O126, and O146 *E. coli*, as well as 20 non-*E. coli* bacterial strains. Moreover, 90 strains belonging to *E. coli* O117 (collected between 1976 and 2005), 77 strains belonging to O126 (collected between 1982 and 2005), and 98 strains belonging to *E. coli* O146 (collected between 1976 and 2006) isolated from humans, animals, food, and water were also included for specificity analysis. PCR assays targeting the *wzx* and *wzy* genes of *E. coli* O126 and O146 were highly specific for their respective serogroups. Cross-reactions did not occur with other *E. coli* strains or with the non-*E. coli* strains tested. Therefore, the *wzx* and *wzy* primers designed in this study could potentially be used in rapid diagnostic assays for *E. coli* serogroups O126 and O146. Interestingly, primers targeting the *wzx* and *wzy* genes of *E. coli* O117 also amplified the *wzx* and *wzy* genes in *E. coli* O107 (Fig. 2). Restriction fragment length polymorphism (RFLP) analyses were performed on the long PCR products of the O-antigen gene clusters of *E. coli* O117 and O107. Since both *E. coli* O117 and O107 had indistinguishable RFLP patterns (data not shown), we speculated that O117 and O107 have identical or at least very similar O-antigen gene cluster sequences. The fact that O117 and O107 antisera have also been found to cross-react supports this notion [30].

In the current study, we determined the DNA sequences of the O-antigen gene clusters of strains representing *E. coli*

Table 4
PCR primers targeting the *wzx* and *wzy* genes of *E. coli* O117, O126, and O146

Target gene	Sequence	Amplicon size (bp)
O117 *wzx*	F: TGTTCTCCACTGCGATCATAGGT R: ACATAGAGTACCCGACACCATCAC	518
O117 *wzy*	F: TGA AAT ACT CGG TTT CAG CAA GAG R: TAG CCA GCA AGG TAT GCT GAA GGA	298
O126 *wzx*	F: TTA GCT CTC GTA GAG GCT GGT GTT R: ATG TCA TTC CTG GGA CGC GAA TGT	925
O126 *wzy*	F: CGC ATT AAA TGG ACC TGA TAA AGC ATC G R: ACT AGC GCA CAT ATC GTT AGC ACG	465
O146 *wzx*	F: AGG GTG ACC ATC AAC ACA CTT GGA R: AGT TCA ATA CTG TCG CAG CTC CTC	640
O146 *wzy*	F: ATT CGG GTA ACG ACC CTG TGT TGA R: AGA CTG CTA ATG CAA GGA ACA TGG	378

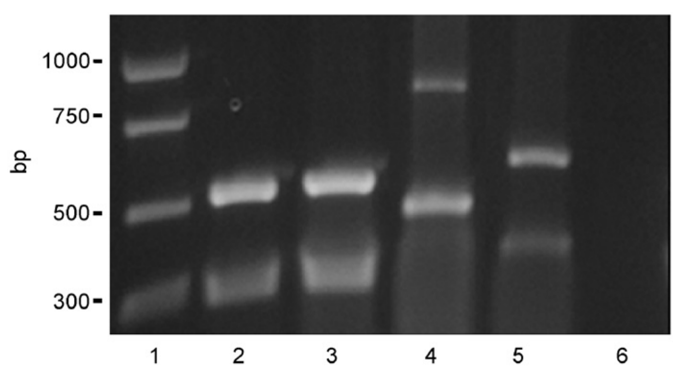

Fig. 2. Results of multiplex PCR assays targeting the *E. coli* O117, O126, and O146 *wzx* and *wzy* genes. Lane 1, Molecular weight markers. Lane 2, PCR products of *E. coli* O117 *wzx* (518 bp) and *wzy* (298 bp) genes. Lane 3, PCR products of *E. coli* O107 *wzx* (518 bp) and *wzy* (198 bp) genes. Lane 4, PCR products of *E. coli* O126 *wzx* (925 bp) and *wzy* (465 bp) genes. Lane 5, PCR products of *E. coli* O146 *wzx* (640 bp) and *wzy* (378 bp) genes. Lane 6, negative control with no template DNA.

serogroups O117, O126, and O146, identified the genes in the O-antigen gene clusters, and developed specific PCR primers for each serogroup targeting the *wzx* and *wzy* genes. Since strains of *E. coli* O117, O126, and O146 are human pathogens, identifying diagnostic markers will help in the clinical diagnosis of diseases caused by these *E. coli* serogroups. Sequence information of the O-antigen gene clusters will also assist in developing microarray-based assays for *E. coli* serotyping. Recently, DNA microarrays were developed for rapid identification of different serogroups of *E. coli* in a single platform [31]. Unique sequences in the O-antigen gene clusters of several *E. coli* serogroups were selected and placed onto the microarrays, and specific signals were generated for each serogroup tested. Sequencing of additional *E. coli* O-antigen gene clusters will allow us to expand the DNA array to contain unique target sequences for all the *E. coli* serogroups.

The O-antigen lipopolysaccharide has been associated with a number of biological phenomena including stress responses [32], swarming motility [33], virulence [34], and flagellum biogenesis [35]. Elucidation of O-antigen gene sequences will provide additional insight on these biological phenomena. In addition, characterization of the O-antigen gene clusters will aid in designing diagnostic markers for specific serogroups and in the development of *E. coli* O-antigen-based vaccines.

Acknowledgments

We are grateful to Drs. John Luchansky and Shu-I Tu (Eastern Regional Research Center, Wyndmoor, PA) for their support. We thank Mr. Paul Pierlott (Eastern Regional Research Center, Wyndmoor, PA) for processing the images. We also thank Dr. Connie Briggs, Dr. David Needleman, and Ms. Laurie Fortis (Eastern Regional Research Center, Wyndmoor, PA) for assisting in sequencing the O-antigen gene clusters of *E. coli* O117, O126, and O146, and Ms. Elizabeth Roberts at the Pennsylvania State University for performing the PCR assays.

References

[1] Smith JL, Fratamico PM. Diarrhea-inducing *Escherichia coli*. In: Fratamico PM, Bhunia AK, Smith JL, editors. Foodborne pathogens microbiology and molecular biology. Norfolk, UK: Caister Academic Press; 2005. p. 357–82.

[2] Aubel D, Darfeuille-Michaud A, Joly B. New adhesive factor (Antigen 8786) on a human enterotoxigenic *Escherichia coli* O117:H4 strain Isolated in Africa. Infect Immun 1991;59:1290–9.

[3] Forestier C, Darfeuille-Michaud A, Wasch E, Rich C, Petat E, Denis F, et al. Adhesive properties of enteropathogenic *Escherichia coli* isolated from infants with acute diarrhea in Africa. Eur J Clin Microbiol Infect Dis 1989;8:979–83.

[4] Hebelka M, Lincoln K, Sandberg T. Sexual acquisition of acute pyelonephritis. Scand J Infect Dis 1993;25:141–3.

[5] Levine MM. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. J Infect Dis 1987;155:377–89.

[6] Yam WC, Robins-Browne RM, Lung ML. Genetic relationships and virulence factors among classical enteropathogenic *Escherichia* serogroup O126 strains. J Med Microbiol 1994;40:229–35.

[7] Cortes C, De la Fuente R, Blanco J, Blanco M, Blanco JE, Dhabi G, et al. Serotypes, virulence genes and intimin types of verotoxin-producing *E. coli* and enteropathogenic *Escherichia coli* isolated from healthy dairy goats in Spain. Vet Microbiol 2005;110:67–76.

[8] Paton AW, Paton JC. Molecular characterization of the locus encoding biosynthesis of the lipopolysaccharide O antigen of *Escherichia coli* serotype O113. Infect Immun 1999;67:5930–7.

[9] Beutin L, Krause G, Zimmermann S, Kaulfuss S, Gleier K. Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period. J Clin Microbiol 2004;42:1099–108.

[10] Beutin L, Geier D, Steinruck H, Zimmermann S, Scheutz F. Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals. J Clin Microbiol 1993;31:2483–8.

[11] Padola NL, Sanz ME, Blanco JE, Blanco M, Blanco J, Etcheverria AI, et al. Serotypes and virulence genes of bovine shigatoxigenic *Escherichia coli* (STEC) isolated from a feedlot in Argentina. Vet Microbiol 2004;100:3–9.

[12] Vettorato MP, Leomil L, Guth BE, Irino K, Pestana de Castro AF. Properties of Shiga toxin-producing *Escherichia coli* (STEC) isolates from sheep in the State of Sao Paulo, Brazil. Vet Microbiol 2003;95:103–9.

[13] Rey J, Blanco JE, Blanco M, Mora A, Dahbi G, Alonso JM, et al. Serotypes, phage types and virulence genes of Shiga-producing *Escherichia coli* isolated from sheep in Spain. Vet Microbiol 2003;94:47–56.

[14] Samuel G, Reeves PR. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res 2003;338:2503–19.

[15] Beutin L, Long Q, Feng L, Wang Q, Krause G, Leomil L, et al. Development of PCR assays targeting the gene involved in synthesis and assembly of the new *Escherichia coli* O174 and O177 O antigens. J Clin Microbiol 2005;43:5143–9.

[16] Feng L, Han W, Wang Q, Bastin DA, Wang L. Characterization of *Escherichia coli* O86 O-antigen gene cluster and identification of O86-specific genes. Vet Microbiol 2005;106:241–8.

[17] Fratamico PM, Briggs CE, Needle D, Chen CY, DebRoy C. Sequence of the *Escherichia coli* O121 O-antigen gene cluster and detection of enterohemorrhagic *Escherichia coli* O121 by PCR amplification of the *wzx* and *wzy* genes. J Clin Microbiol 2003;41:3379–83.

[18] Shao J, Li M, Jia Q, Lu Y, Wang PG. Sequence of *Escherichia coli* O128 antigen biosynthesis cluster and functional identification of an alpha-1,2-fucosyltransferase. FEBS Lett 2003;553:99–103.

[19] Wang L, Curd H, Qu W, Reeves PR. Sequencing of *Escherichia coli* O111 O-antigen gene cluster and identification of O111-specific genes. J Clin Microbiol 1998;36:3182–7.

[20] Wang L, Qu W, Reeves PE. Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships. Infect Immun 2001;68:6823–930.

[21] Wang L, Huskic S, Cisterne A, Rothemund D, Reeves PR. The O-antigen gene cluster of *E. coli* O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. J Bacteriol 2002;184:2620–5.

[22] Ørskov I, Ørskov F, Jann B, Jann K. Serology, chemistry and genetics of O and K antigens of *Escherichia*. Bacteriol Rev 1977;41:667–710.

[23] Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics 2001;17:849–50.

[24] Leslie MR, Parolis H, Parolis LA. The structure of the O-specific polysaccharide of *Escherichia coli* O117:K98:H4. Carbohydr Res 2000;323:103–10.

[25] Kang YB, Yang YH, Lee KW, Lee SG, Sohng JK, Lee HC, et al. Preparative synthesis of dTDP-L-rhamnose through combined enzymatic pathways. Biotechnol Bioeng 2006;93:21–7.

[26] Fratamico PM, DebRoy C, Strobaugh Jr TP, Chen CY. DNA sequence of the *Escherichia coli* O103 O antigen gene cluster and

detection of enterohemorrhagic *Escherichia coli* O103 by PCR amplification of the *wzx* and *wzy* genes. Can J Microbiol 2005;51(6):515–22.

[27] Thoden JB, Henderson JM, Fridovich-Keil JL, Holden HM. Structural analysis of the Y299C mutant of *Escherichia coli* UDP-galactose 4-epimerase, Teaching an old dog new tricks. J Biol Chem 2002;277:27528–34.

[28] Metzger M, Bellemann P, Bugert P, Geider K. Genetics of galactose metabolism of *Erwinia amylovora* and its influence on polysaccharide synthesis and virulence of the fire blight pathogen. J Bacteriol 1994;176:450–9.

[29] Larsson EA, Urbina F, Yang Z, Weintraub A, Widmalm G. Structural and immunochemical relationship between the O-antigenic polysaccharides from the enteroaggregative *Escherichia coli* strain 396/C-1 and *Escherichia coli* O126. Carbohydr Res 2004;339:1491–6.

[30] Gamage SD, McGannon CM, Weiss AA. *Escherichia coli* serogroup O107/O117 lipopolysaccharide binds and neutralizes Shiga toxin 2. J Bacteriol 2004;186:5506–12.

[31] Liu Y, Fratamico PM. *Escherichia coli* O antigen typing using DNA microarrays. Mol Cellular Probes 2006;20:239–44.

[32] Vines ED, Marolda CL, Balachandran A, Valvano MA. Defective O-antigen polymerization in tolA and pal mutants of *Escherichia coli* in response to extracytoplasmic stress. J Bacteriol 2005;187:3359–68.

[33] Toguchi A, Siano M, Burkart M, Harshey RM. Genetics of swarming motility in *Salmonella enterica* serovar Typhimurium: critical role for lipopolysaccharide. J Bacteriol 2000;182:6308–21.

[34] Canals R, Jimenez N, Vilches S, Regue M, Merino S, Tomas JM. The UDP *N*-acetylgalactosamine 4-epimerase gene is essential for mesophilic *Aeromonas hydrophila* serotype O34 virulence. Infect Immun 2006;74:537–48.

[35] Abeyrathne PD, Daniels C, Poon KK, Matewish MJ, Lam JS. Functional characterization of WaaL, a ligase associated with linking O-antigen polysaccharide to the core of *Pseudomonas aeruginosa* lipopolysaccharide. J Bacteriol 2005;187:3002–12.